# Survival in a quasi-death process

Erik A. van Doorn

Department of Applied Mathematics
University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
E-mail: e.a.vandoorn@utwente.nl

2 January 2007

**Abstract.** We consider a Markov chain in continuous time with an absorbing coffin state and a finite set $S$ of transient states. When $S$ is irreducible the limiting distribution of the chain as $t \to \infty$, conditional on survival up to time $t$, is known to equal the (unique) quasi-stationary distribution of the chain. We address the problem of generalizing this result to a setting in which $S$ may be reducible, and obtain a complete solution if the eigenvalue with maximal real part of the generator of the (sub)Markov chain on $S$ has multiplicity one. The result is applied to pure death processes and, more generally, to quasi-death processes.

*Keywords and phrases:* absorbing Markov chain, death process, limiting conditional distribution, quasi-stationary distribution, survival-time distribution

*2000 Mathematics Subject Classification:* Primary 60J27

# 1 Introduction

In the interesting papers [2] and [3] Aalen and Gjessing provide a new explanation for the shape of hazard rate functions in survival analysis. They propose to model survival times as sojourn times of stochastic processes in a set $S$ of transient states until they escape from $S$ to an absorbing coffin state. This "process point of view" entails that (in the words of Aalen and Gjessing) "the shape of the hazard rate is created in a balance between two forces: the attraction of the absorbing state and the general diffusion within the transient space". As a result the shape of the hazard rate is determined by the interaction of the initial distribution and the distribution over $S$ known as the *quasi-stationary distribution* of the process. Similar ideas have been put forward independently by Steinsaltz and Evans [15].

Aalen and Gjessing discuss several examples of relevant stochastic processes, including finite-state Markov chains with an absorbing state, the setting of the present paper. A survival-time distribution in this setting is known as a *phase-type distribution* (see, for example, Aalen [1]). In their analysis and examples Aalen and Gjessing restrict themselves to chains for which the set $S$ of transient states constitutes a single class, arguing that "irreducibility is important when considering quasistationary distributions". As we shall see, however, there are no compelling technical reasons for imposing this restriction. Moreover, in [3, Section 8] Aalen and Gjessing allude to a bottle-neck phenomenon that may occur when $S$ is reducible, making it even desirable to investigate what happens in this case. We note that Proposition 1 in [15], while formulated quite generally, seems to be entirely correct only if one assumes $S$ to be irreducible.

From a modelling point of view there is another argument for extending the analysis to reducible sets $S$. Namely, if the status of an individual before evanescence is represented by the state of a transient Markov chain, it seems reasonable to allow for the possibility that some transitions are irreversible, reflecting the fact that some real-life processes such as *ageing* are irreversible.

The main aim of the present paper is to provide the tools for hazard rate analysis, by characterizing survival-time distributions and identifying limiting conditional distributions and quasi-stationary distributions, in the setting of finite Markov chains with an absorbing state and a transient space $S$ that may be reducible. In Section 2 we present some general results, which are applied in Section 3 to pure death processes. The latter results are then generalized in Section 4 to *quasi-death processes*, which may be viewed as death processes in which the sojourn time in each state has a phase-type distribution.

## 2 Absorbing Markov chains

Consider a continuous-time Markov chain $\mathcal{X} := \{X(t),\ t \geq 0\}$ on a state space $\{0\} \cup S$ consisting of an absorbing state 0 and a finite set of transient states $S := \{1, 2, \ldots, n\}$. The generator of $\mathcal{X}$ then takes the form

$$\begin{pmatrix} 0 & \mathbf{0} \\ \boldsymbol{q}^T & Q \end{pmatrix}, \tag{1}$$

where

$$\boldsymbol{q} = -\mathbf{1}Q^T > \mathbf{0}. \tag{2}$$

Here $\mathbf{0}$ and $\mathbf{1}$ are row vectors of zeros and ones, respectively, superscript $^T$ denotes transpose, and strict inequality for vectors indicates strict inequality in at least one component. Since all states in $S$ are transient, state 0 is accessible from any state in $S$. Hence, whichever the initial state, the process will eventually escape from $S$ into the absorbing state 0 with probability one.

We write $\mathbb{P}_i(.)$ for the probability measure of the process when $X(0) = i$, and let $\mathbb{P}_{\boldsymbol{w}}(.) := \sum_i w_i \mathbb{P}_i(.)$ for any vector $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$ representing a distribution over $S$. Also, $P_{ij}(.) := \mathbb{P}_i(X(.) = j)$. It is easy to verify (see, for example, Kijima [8, Section 4.6]) that the matrix $P(t) := (P_{ij}(t),\ i, j \in S)$ satisfies

$$P(t) = e^{Qt} := \sum_{k=0}^{\infty} \frac{Q^k}{k!} t^k, \quad t \geq 0.$$

By $T := \sup\{t \geq 0 : X(t) \in S\}$ we denote the *survival time* (or *absorption time*) of $\mathcal{X}$, the random variable representing the time at which escape from

$S$ occurs. In what follows we are interested in the limiting distribution of the residual survival time conditional on survival up to time $t$, that is,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(T \le t+s \,|\, T > t), \quad s \ge 0, \tag{3}$$

and in the limiting distribution of $X(t)$ conditional on survival up to time $t$, that is,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t), \quad j \in S, \tag{4}$$

where $\boldsymbol{w}$ is any initial distribution over $S$.

Let us first suppose that $S$ is irreducible, that is, constitutes a single communicating class. In this case $Q$ has a unique eigenvalue with maximal real part, which we denote by $-\alpha$. It is well known (see, for example, Seneta [14, Theorem 2.6]) that $\alpha$ is real and positive, and that the associated left and right eigenvectors $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ can be chosen strictly positive componentwise. It will also be convenient to normalize $\boldsymbol{u}$ and $\boldsymbol{v}$ such that

$$\boldsymbol{u}\mathbf{1}^T = 1 \;\text{ and }\; \boldsymbol{u}\boldsymbol{v}^T = 1. \tag{5}$$

It then follows (see Mandl [12]) that the transition probabilities $P_{ij}(t)$ satisfy

$$\lim_{t\to\infty} e^{\alpha t} P_{ij}(t) = v_i u_j, \quad i, j \in S, \tag{6}$$

which explains why $\alpha$ is often referred to as the *decay parameter* of $\mathcal{X}$. We shall show later (Theorem 4) that (6) actually holds true in a more general setting.

Since $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$, we have $\boldsymbol{u}Q^k = (-\alpha)^k\boldsymbol{u}$ for all $k$, and hence

$$\boldsymbol{u}P(t) = \sum_{k=0}^{\infty} \frac{\boldsymbol{u}Q^k}{k!} t^k = e^{-\alpha t}\boldsymbol{u}, \quad t \ge 0,$$

that is

$$\mathbb{P}_{\boldsymbol{u}}(X(t) = j) = e^{-\alpha t} u_j, \quad j \in S,\; t \ge 0. \tag{7}$$

Since $\mathbb{P}_{\boldsymbol{u}}(T > t) = \mathbb{P}_{\boldsymbol{u}}(X(t) \in S) = e^{-\alpha t}$, it follows that for all $t \ge 0$

$$\mathbb{P}_{\boldsymbol{u}}(T > t+s \,|\, T > t) = e^{-\alpha s}, \quad s \ge 0. \tag{8}$$

Moreover, $\boldsymbol{u}$ is a *quasi-stationary distribution* of $\mathcal{X}$ in the sense that for all $t \geq 0$

$$\mathbb{P}_{\boldsymbol{u}}(X(t) = j \,|\, T > t) = u_j, \quad j \in S, \tag{9}$$

that is, the distribution of $X(t)$ conditional on absorption not yet having taken place at time $t$ is constant over $t$ when $\boldsymbol{u}$ is the initial distribution. Darroch and Seneta [5] have shown that similar results hold true in the limit as $t \to \infty$ when the initial distribution differs from $\boldsymbol{u}$. Namely, for any initial distribution $\boldsymbol{w}$ one has

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(T > t + s \,|\, T > t) = e^{-\alpha s}, \quad s \geq 0, \tag{10}$$

and

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t) = u_j, \quad j \in S. \tag{11}$$

So when all states in $S$ communicate the limits (3) and (4) are determined by the largest eigenvalue of $Q$ and the corresponding left eigenvector.

This result can be generalized, at least in principle, to a setting in which $S$ consists of more than one class. Indeed, suppose that $S$ consists of communicating classes $S_1, S_2, \ldots, S_c$, and let $Q_k$ be the submatrix of $Q$ corresponding to the states in $S_k$. Obviously, the set of eigenvalues of $Q$ is precisely the union of the sets of eigenvalues of the individual $Q_k$'s. So, if we denote the (unique) eigenvalue with maximal real part of $Q_k$ by $-\alpha_k$ (so that $\alpha_k$ is real and positive), and let $\alpha := \min_k \alpha_k$, then $-\alpha$ is the eigenvalue of $Q$ with maximal real part. Evidently, $-\alpha$ may be degenerate, but we will restrict ourselves to settings in which $-\alpha$ has algebraic (and hence geometric) multiplicity one. Under this condition then there exist, up to constant factors, unique left and right eigenvectors $\boldsymbol{u}$ and $\boldsymbol{v}$ corresponding to $-\alpha$. It follows from Theorem I* of Debreu and Herstein [6] (by an argument similar to the proof of [14, Theorem 2.6]) that we may choose $\boldsymbol{u} > \mathbf{0}$, $\boldsymbol{v} > \mathbf{0}$ and $\boldsymbol{u}\mathbf{1}^T = 1$, but $\boldsymbol{u}$ and $\boldsymbol{v}$ are not necessarily positive componentwise.

In the present setting (7), and hence (8) and (9), retain their validity. Letting

$$a(\alpha) := \arg\min_k \alpha_k, \tag{12}$$

we note that $S_{a(\alpha)}$ must be accessible from $\boldsymbol{u}$ (that is, accessible from a state $i$ such that $u_i > 0$). Indeed, $\alpha$ having multiplicity one, the opposite would imply that (8) cannot be true. It is well known that $\mathbb{P}_{\boldsymbol{u}}(X(t) = j) > 0$ for all $t > 0$ if and only if $j$ is accessible from $\boldsymbol{u}$, so it follows from (7) that we must actually have $u_j > 0$ for all states $j$ that are accessible from $\boldsymbol{u}$, and in particular for all states $j$ that are accessible from $S_{a(\alpha)}$. On the other hand, $\boldsymbol{u}$ being the unique solution of the system $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$, we must have $u_j = 0$ if $j$ is *not* accessible from $S_{a(\alpha)}$. For it is easily seen that we can determine $\boldsymbol{u}$ by first solving the eigenvector problem in the restricted setting of states that are accessible from $S_{a(\alpha)}$, and subsequently putting $u_j = 0$ whenever $j$ is not accessible from $S_{a(\alpha)}$. So $u_j > 0$ if and only if state $j$ is accessible from $S_{a(\alpha)}$. The counterpart of (7) for the right eigenvector $\boldsymbol{v}$ is the relation

$$\sum_{j\in S} P_{ij}(t)v_j = e^{-\alpha t}v_i, \quad i \in S, \tag{13}$$

which may be used in a similar way to show that $v_i > 0$ if and only if $S_{a(\alpha)}$ is accessible from $i$. It follows in particular that both $u_j > 0$ *and* $v_j > 0$ if (and only if) $j \in S_{a(\alpha)}$, so that $\boldsymbol{v}$ may be normalized such that $\boldsymbol{u}\boldsymbol{v}^T = 1$. We summarize our findings in the next theorem.

**Theorem 1** If $-\alpha$, the eigenvalue of $Q$ with maximal real part, has multiplicity one, then there are unique nonnegative vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfying $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$, $Q\boldsymbol{v}^T = -\alpha\boldsymbol{v}^T$, $\boldsymbol{u}\mathbf{1}^T = 1$, and $\boldsymbol{u}\boldsymbol{v}^T = 1$. The $i$th component of $\boldsymbol{u}$ is positive if and only if state $i$ is accessible from $S_{a(\alpha)}$, whereas the $i$th component of $\boldsymbol{v}$ is positive if and only if $S_{a(\alpha)}$ is accessible from state $i$.

The vector $\boldsymbol{u}$ does not necessarily constitute the only quasi-stationary distribution of the process $\mathcal{X}$, that is, the only initial distribution satisfying (9) for all $t \geq 0$. However, we can achieve uniqueness if we restrict ourselves to initial distributions from which $S_{a(\alpha)}$ is accessible. To prove this statement we need the following invariance result.

**Lemma 2** If the initial distribution $\boldsymbol{w}$ is such that $S_{a(\alpha)}$ is accessible, and satisfies $\boldsymbol{w}Q = x\boldsymbol{w}$ for some $x < 0$, then $x = -\alpha$ and $\boldsymbol{w} = \boldsymbol{u}$.

**Proof** When the initial distribution $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$ is a left eigenvector corresponding to the eigenvalue $x$, then, by an argument similar to the one leading to (7), we have

$$\mathbb{P}_{\boldsymbol{w}}(X(t) = j) = e^{xt} w_j, \quad j \in S, \ t \geq 0.$$

It follows that $w_j > 0$ for all states $j$ that are accessible from $\boldsymbol{w}$. So, if $S_{a(\alpha)}$ is accessible from $\boldsymbol{w}$, then $w_j > 0$ for all $j \in S_{a(\alpha)}$. Hence, by Theorem 1, $\boldsymbol{w}\boldsymbol{v}^T > 0$. Since $\boldsymbol{w}Q = x\boldsymbol{w}$ implies $x\boldsymbol{w}\boldsymbol{v}^T = \boldsymbol{w}Q\boldsymbol{v}^T = -\alpha\boldsymbol{w}\boldsymbol{v}^T$, we must have $x = -\alpha$, and hence $\boldsymbol{w} = \boldsymbol{u}$. $\square$

We can now copy the arguments in [5] (in which a similar invariance result is implicitly used) and conclude the following.

**Theorem 3** If $-\alpha$, the eigenvalue of $Q$ with maximal real part, has multiplicity one, then $\mathcal{X}$ has a unique quasi-stationary distribution $\boldsymbol{u}$ from which $S_{a(\alpha)}$ is accessible. The vector $\boldsymbol{u}$ is the (unique, nonnegative) solution of the system $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$.

To determine the limits (10) and (11) in the general setting at hand we need the announced generalization of (6). Its proof is similar to the proof of Theorem 1 in [12], but since this reference is in Russian we sketch the argument.

**Theorem 4** If $-\alpha$, the eigenvalue of $Q$ with maximal real part, has multiplicity one then

$$\lim_{t \to \infty} e^{\alpha t} P(t) = \boldsymbol{v}^T \boldsymbol{u}, \tag{14}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are the eigenvectors defined in Theorem 1.

**Proof** With $J = (J_{ij})$ denoting the Jordan canonical form of $Q$, there exists a nonsingular matrix $S = (S_{ij})$ such that $Q = SJS^{-1}$, and hence

$$P(t) = e^{tQ} = Se^{tJ}S^{-1}, \quad t \geq 0.$$

Since $J_{11} = -\alpha$, while $J_{1j} = J_{j1} = 0$ if $j \neq 1$, it follows that

$$P_{ij}(t) = e^{-\alpha t} S_{i1}(S^{-1})_{1j} + o\left(e^{-\alpha t}\right) \quad \text{as} \quad t \to \infty, \quad i, j \in S,$$

and hence

$$\lim_{t\to\infty} e^{\alpha t}P(t) = \boldsymbol{s}^T\boldsymbol{t},$$

where $\boldsymbol{s}^T$ denotes the first column of $S$ and $\boldsymbol{t}$ the first row of $S^{-1}$. Since $QS = SJ$ we must have $Q\boldsymbol{s}^T = -\alpha\boldsymbol{s}^T$, so we can normalize $\boldsymbol{s}$ such that $\boldsymbol{s} = \boldsymbol{v}$. Moreover, by the Markov property,

$$e^{-\alpha s}\boldsymbol{v}^T\boldsymbol{t} = e^{-\alpha s}\lim_{t\to\infty} e^{\alpha(t+s)}P(t+s) = \boldsymbol{v}^T\boldsymbol{t}P(s).$$

Pre-multplying this relation by $\boldsymbol{u}$ we obtain $e^{-\alpha s}\boldsymbol{t} = \boldsymbol{t}P(s)$. Subsequently taking derivatives with respect to $s$, and letting $s \downarrow 0$ yields $-\alpha\boldsymbol{t} = \boldsymbol{t}Q$. Finally, since $\boldsymbol{t}\boldsymbol{v}^T = \boldsymbol{t}\boldsymbol{s}^T = 1$, we must have $\boldsymbol{t} = \boldsymbol{u}$. □

We can now copy the argument in [12] or [5] to conclude the following.

**Theorem 5** If $-\alpha$, the eigenvalue of $Q$ with maximal real part, has multiplicity one, and the initial distribution $\boldsymbol{w}$ is such that $S_{a(\alpha)}$ is accessible, then the limits (3) and (4) exist and are given by (10) and (11), respectively, where $\boldsymbol{u}$ is the unique quasi-stationary distribution from which $S_{a(\alpha)}$ is accessible.

**Remark** The results in [12] and [5] constitute the continuous-time counterparts of results obtained in [11] and [4], respectively, in a discrete-time setting. The latter results have been generalized (in a more abstract, but still discrete, setting) by Lindqvist [10]. An alternative approach towards proving Theorem 5 would be to take Lindqvist results (in particular [10, Theorem 5.8]) as a starting point and prove their analogues in a continuous setting. In this way an even more general statement would result (allowing a degenerate eigenvalue $-\alpha$ under certain conditions), but at the cost of a more elaborate notation and formulation.

The fact that the limiting distribution of the residual survival time exists and is exponentially distributed has been observed by Kalpakam [7] and Li and Cao [9] in a somewhat more general setting, namely when the Laplace transform of the survival-time distribution is a rational function (cf. [13]).

In what follows we are interested in particular in properties of the left eigenvector $\boldsymbol{u}$ that are determined by structural properties of $Q$. To set the stage we first look more closely into the simple multi-class setting of a pure death process in the next section, and then generalize our results to quasi-death processes in Section 4.

## 3 Pure death processes

Let us assume that the Markov chain $\mathcal{X} = \{X(t),\ t \geq 0\}$ of the previous section is a pure death process with death rate $\mu_i$ in state $i \in S$, so that the matrix $Q$ of (1) is given by

$$Q = \begin{pmatrix} -\mu_1 & 0 & 0 & . & . & . & 0 & 0 & 0 \\ \mu_2 & -\mu_2 & 0 & . & . & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & 0 & \mu_n & -\mu_n \end{pmatrix}. \tag{15}$$

Evidently, the classes of $S$ now consist of single states, so, maintaining the notation of the previous section, we let $S_k = \{k\}$, and find that $\alpha_k = \mu_k$ and

$$\alpha = \mu := \min_{i \in S} \mu_i. \tag{16}$$

As before, we assume that $\mu$ is a nondegenerate eigenvalue of $Q$, whence

$$a := \arg\min_{i \in S} \mu_i \tag{17}$$

is uniquely defined. It is clear that an initial distribution $\boldsymbol{w}$ satisfies the requirements of Theorem 5 if and only if $\boldsymbol{w}$ has support in the set of states $\{a, a+1, \ldots, n\}$. Theorem 5 therefore implies the following, where an empty product denotes unity.

**Theorem 6** Let $\mathcal{X}$ is a pure death process with death rate $\mu_i$ in state $i \in S$, and a unique state $a$ such that $\mu_a = \min_{i \in S} \mu_i$. If the initial distribution $\boldsymbol{w}$ is supported by at least one state $i \geq a$, then

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(T > t + s \,|\, T > t) = e^{-\mu s}, \quad s \geq 0. \tag{18}$$

and

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t) = u_j, \quad j \in S, \tag{19}$$

where $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ is the (unique) quasi-stationary distribution of $\mathcal{X}$ from which $S_{a(\alpha)}$ is accessible, and given by

$$u_j = \begin{cases} \dfrac{\mu}{\mu_j} \displaystyle\prod_{i=1}^{j-1} \left(1 - \frac{\mu}{\mu_i}\right), & j < a \\ \displaystyle\prod_{i=1}^{a-1} \left(1 - \frac{\mu}{\mu_i}\right), & j = a \\ 0, & j > a. \end{cases} \tag{20}$$

**Proof** By Theorems 3 and 5 we have to show that the vector $\boldsymbol{u}$ satisfies $\boldsymbol{u}Q = -\mu\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$. It is a routine exercise to verify these properties. □

**Example** The quasi-stationary distribution of the death process on $S = \{0, 1, 2\}$ is given by

$$\boldsymbol{u} = (u_1, u_2) = \begin{cases} \left(\dfrac{\mu_2}{\mu_1}, 1 - \dfrac{\mu_2}{\mu_1}\right) & \text{if } \mu_2 < \mu_1 \\ (1, 0) & \text{if } \mu_1 < \mu_2. \end{cases} \tag{21}$$

In view of Theorem 5 we conclude that state 1 is a *bottle-neck* state when $\mu_1 < \mu_2$, in the sense that the process is almost surely in state 1 if, after a long time, absorption has not yet occurred, whatever the initial distribution. This is an example of the phenomenon alluded to by Aalen and Gjessing in [3, Section 8]. Note that $(1, 0)$ is also a quasi-stationary distribution if $\mu_2 < \mu_1$, but one from which state 2 is not accessible. So it is a limiting conditional distribution only if $\mathbb{P}(X(0) = 2) = 0$. □

As an aside we remark that the survival time in any birth-death process can be represented by the survival time in a pure death process with the same number of states (see, for example, Aalen [1]). Evidently, the quasi-stationary distributions of the two processes will be different in general.

## 4 Quasi-death processes

The absorbing continuous-time Markov chain $\mathcal{X} := \{X(t),\ t \geq 0\}$ of Section 2 is a *quasi-death process* if $S = \{(\ell, j) \,|\, \ell = 1, 2, \ldots, L,\ j = 1, 2, \ldots, J_\ell\}$ and $Q$ takes the block-partitioned form

$$Q = \begin{pmatrix} Q_1 & 0 & 0 & . & . & 0 & 0 & 0 \\ M_2 & Q_2 & 0 & . & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 0 & M_L & Q_L \end{pmatrix}, \tag{22}$$

where $Q_\ell$ and $M_\ell$ are nonzero matrices of dimension $J_\ell \times J_\ell$, and $J_\ell \times J_{\ell-1}$, respectively. We write $X(t) = (L(t), J(t))$ and call $L(t)$ the *level* and $J(t)$ the *phase* of the process at time $t < T$. Throughout this section we assume that $S_\ell := \{(\ell, j) \,|\, j = 1, 2, \ldots, J_\ell\}$ is a communicating class for each level $\ell$. Moreover, we suppose

$$\mathbf{1}M_\ell^T + \mathbf{1}Q_\ell^T = \mathbf{0}, \quad \ell = 2, 3, \ldots, L, \tag{23}$$

and, to be consistent with (2),

$$\boldsymbol{q}_1 := -\mathbf{1}Q_1^T > \mathbf{0}. \tag{24}$$

Hence, with probability one and for any initial state $(\ell, i)$, the function $L(t),\ 0 \leq t < T$, will be a step function with downward jumps of size one, and the process will eventually escape from $S$, via a state at level 1, to the absorbing state 0. Extending the notation introduced in Section 2 we write

$$\mathbb{P}_{\boldsymbol{w}_\ell}(.) := \sum_{i=1}^{J_\ell} w_{\ell i}\mathbb{P}_{(\ell,i)}(.)$$

for any distribution $\boldsymbol{w}_\ell = (w_{\ell 1}, w_{\ell 2}, \ldots, w_{\ell J_\ell})$ over $S_\ell$.

Evidently, if $J_\ell = 1$ for all levels $\ell$ then we are in the setting of the simple death process of the previous section with death rate $\mu_1 := \boldsymbol{q}_1$ in state 1 and $\mu_\ell := M_\ell$ in state $\ell > 1$. On the other hand, if the initial distribution concentrates all mass on the first level, we are basically dealing with a Markov chain taking values in the set $\{0\} \cup S_1$, with 0 an absorbing state and $S_1$ a single communicating class, a setting discussed in the beginning of Section 2. In the general setting at hand we must apply Theorems 3 and 5, but, as we shall see, we can reduce the amount of computation by exploiting the structure of $Q$.

We denote the (unique) eigenvalue of $Q_\ell$ with maximal real part by $-\alpha_\ell$, and the associated left and right eigenvectors by $\boldsymbol{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell J_\ell})$ and $\boldsymbol{y}_\ell = (y_{\ell 1}, y_{\ell 2}, \ldots, y_{\ell J_\ell})$, respectively. As noted before, $\alpha_\ell$ is real and positive, and $\boldsymbol{x}_\ell$ and $\boldsymbol{y}_\ell$ can be chosen strictly positive componentwise and such that

$$\boldsymbol{x}_\ell \mathbf{1}^T = 1 \quad \text{and} \quad \boldsymbol{x}_\ell \boldsymbol{y}_\ell^T = 1. \tag{25}$$

In analogy to (6) we have

$$\lim_{t\to\infty} e^{\alpha_\ell t} P_{(\ell,i),(\ell,j)}(t) = y_{\ell i} x_{\ell j}, \quad i, j = 1, 2, \ldots, J_\ell, \tag{26}$$

for each level $\ell$, so we will refer to $\alpha_\ell$ as the *decay parameter of $\mathcal{X}$ in $S_\ell$*. Moreover, the vector $\boldsymbol{x}_\ell$ can be interpreted as the *quasi-stationary distribution of $\mathcal{X}$ in $S_\ell$*, in the sense that

$$\mathbb{P}_{\boldsymbol{u}_\ell}(X(t) = (\ell, j) \,|\, T_\ell > t) = x_{\ell j}, \quad t \geq 0, \; j = 1, 2, \ldots, J_\ell, \tag{27}$$

where $T_\ell$ denotes the sojourn time of $\mathcal{X}$ in $S_\ell$, while

$$\mathbb{P}_{\boldsymbol{u}_\ell}(T_\ell > t) = e^{-\alpha_\ell t}, \quad t \geq 0. \tag{28}$$

If the initial distribution concentrates all mass in $S_\ell$ (and is represented by the vector $\boldsymbol{w}_\ell = (w_{\ell 1}, w_{\ell 2}, \ldots, w_{\ell J_\ell})$, say) but is otherwise arbitrary, then, by the results of Darroch and Seneta [5] mentioned in Section 2,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}_\ell}(X(t) = (\ell, j) \,|\, T_\ell > t) = x_{\ell j}, \quad j = 1, 2, \ldots, J_\ell, \tag{29}$$

and

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}_\ell}(T_\ell > t + s \,|\, T_\ell > t) = e^{-\alpha_\ell s}, \quad s \geq 0. \tag{30}$$

Now turning to a general initial distribution $\boldsymbol{w} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_L)$, where $\boldsymbol{w}_\ell = (w_{\ell 1}, w_{\ell 2}, \ldots, w_{\ell J_\ell})$ for $\ell = 1, 2, \ldots, L$, Theorem 5 tells us that the limiting distribution of the residual survival time in $S = \cup_\ell S_\ell$ is exponentially distributed with parameter $\alpha = \min_k \alpha_k$. As regards the limiting distribution of $X(t)$ conditional on survival in $S$ up to time $t$, we can finally state the following generalization of Theorem 6.

**Theorem 7** Let $\mathcal{X}$ be a quasi-death process for which $Q$ takes the form (22), and which has a unique level $a$ such that $\alpha_a = \min_\ell \alpha_\ell$. If the initial distribution $\boldsymbol{w}$ is supported by at least one state in the set $\cup_{\ell \geq a} S_\ell$, then

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = (\ell, j) \,|\, T > t) = u_{\ell j},\ j = 1, 2, \ldots, J_\ell,\ \ell = 1, 2, \ldots, L, \quad (31)$$

where $\boldsymbol{u}_\ell := (u_{\ell 1}, u_{\ell 2}, \ldots, u_{\ell J_\ell})$ satisfies $\boldsymbol{u}_\ell = \mathbf{0}$ if $\ell > a$, and $\boldsymbol{u}_a = c\boldsymbol{x}_a$, with $\boldsymbol{x}_a$ the (unique and strictly positive) solution of

$$\boldsymbol{x}_a Q_a = -\alpha \boldsymbol{x}_a, \quad \boldsymbol{x}_a \mathbf{1}^T = 1; \quad (32)$$

for $\ell < a$, $\boldsymbol{u}_\ell$ is recursively defined by

$$\boldsymbol{u}_\ell = -\boldsymbol{u}_{\ell+1} M_{\ell+1} (Q_\ell + \alpha I)^{-1}. \quad (33)$$

Here $I$ is an identity matrix of appropriate dimensions and $c > 0$ is such that $\boldsymbol{u}\mathbf{1}^T = 1$, where $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_L)$.

**Proof** Since, for all $\ell \neq a$, the matrix $Q_\ell + \alpha I$ has largest eigenvalue $-(\alpha_\ell - \alpha) < 0$, it follows from [14, Theorem 2.6(g)] that $-(Q_\ell + \alpha I)^{-1}$ exists and has strictly positive components. So, by induction, $\boldsymbol{u}_\ell$ is positive componentwise for $\ell \leq a$. It follows easily that the vector $\boldsymbol{u}$ satisfies the requirements of Theorem 3. □

# References


[1] Aalen, O.O. (1995) Phase type distributions in survival analysis. *Scand. J. Statist.* **22**, 447-463.

[2] Aalen, O.O. and Gjessing, H.K. (2001) Understanding the shape of the hazard rate: a process point of view. *Statist. Sci.* **16**, 1-22.

[3] Aalen, O.O. and Gjessing, H.K. (2003) A look behind survival data: underlying processes and quasi-stationarity. In: Lindqvist, B.H. and Doksum, K.A. (Eds.), *Mathematical and Statistical Methods in Reliability.* World Scientific Publishing, Singapore, pp. 221-234.

[4] Darroch, J.N. and Seneta, E. (1964) On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Probab.* **2**, 88-100.

[5] Darroch, J.N. and Seneta, E. (1967) On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Probab.* **4**, 192-196.

[6] Debreu, G. and Herstein, I.N. (1953) Nonnegative square matrices. *Econometrica* **21**, 597-607.

[7] Kapalkam, S. (1993) On the quasi-stationary distribution of the residual lifetime. *IEEE Trans. Reliab.* **42**, 623-624.

[8] Kijima, M. *Markov Processes for Stochastic Modeling*. Chapman & Hall, London, 1997.

[9] Li, W. and Cao, J. (1993) The limiting distributions of the residual lifetimes of a Markov repairable system. *Reliab. Eng. System Safety* **41**, 103-105.

[10] Lindqvist, B.H. (1989) Asymptotic properties of powers of nonnegative matrices, with applications. *Linear Algebra Appl.* **114/115**, 555-588.

[11] Mandl, P. (1959) On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov chain. *Časopis Pěst. Mat.* **84,** 140-149 (in Russian).

[12] Mandl, P. (1960) On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov process. *Časopis Pěst. Mat.* **85,** 448-455 (in Russian).

[13] O'Cinneide, C.A. (1990) Characterization of phase type distributions. *Stoch. Models* **6**, 1-57.

[14] Seneta, E. *Non-negative Matrices and Markov Chains,* rev. ed., Springer, New York, 1981.

[15] Steinsaltz, D. and Evans, S.N. (2004) Markov mortality models: implications of quasistationarity and varying initial distributions. *Theoret. Population Biol.* **65**, 319-337.